FILTER OPTIMIZATION AND COMPLEXITY REDUCTION FOR VIDEO CODING USING GRAPH-BASED TRANSFORMS

Eduardo Martínez-Enríquez, Fernando Díaz-de-María, Jesús Cid-Sueiro[1] and Antonio Ortega[2]

[1] Department of Signal Theory and Communications. Universidad Carlos III, Leganés (Madrid), Spain

[2]Department of Electrical Engineering. University of Southern California, Los Angeles, California, USA

Motivation

• Directional transforms avoid filtering across large discontinuities→ Smaller high frequency coefficients in those locations.

• **Previous work [1]: Video encoder based on 3-D directional transforms.**

Related Work: Secker and Taubman, 2003; Popescu and Bottreau, 2001.

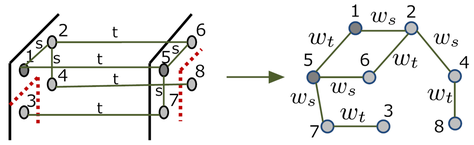
• **Two mayor drawbacks of [1]:**
• Fixed weights on the graph → Fixed prediction filters.
• Low complexity version of the transform depends on the video content.

Key Novelties

Introduction to the Transform:

• Describe the video sequence as a **weighted graph** of connected pixels.

• Apply the lifting transform on this graph.



• **Graph captures spatio-temporal correlation → spatio-temporal filtering operations.**

• **Non separable approach**, against common Wavelet-based video coders (t+s).

Key Novelties:

• Weights of the graph will influence the efficiency of the transform → Contribution : **Find the weights that minimize the detail coefficients energy in a graph-based lifting transform, improving the energy compaction ability of the transform.**

• **Reduce the computational complexity of the transform using a distributed Update/Prediction splitting method.**

Lifting Transforms on Graphs

Lifting Transform:

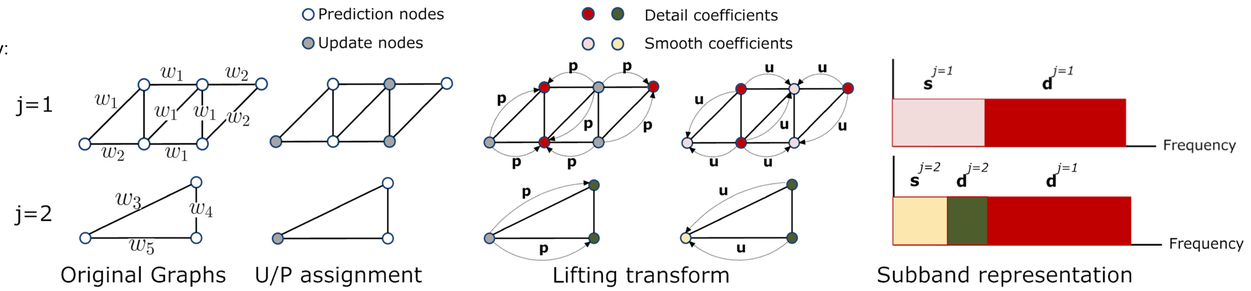
• To perform the transform and ensure its invertibility:

• Update (U_j)- Prediction (P_j) splitting (we use *Weighted maximum-cut* criterion).

• Predict (**p**) and update (**u**) filters design.

$$d_{m,j} = s_{m,j-1} - \sum_{h \in \mathcal{U}_j} \mathbf{p}_{m,j}(h) s_{h,j-1}$$

$$s_{n,j} = s_{n,j-1} + \sum_{l \in \mathcal{P}_j} \mathbf{u}_{n,j}(l) d_{l,j}.$$

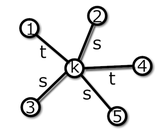


Optimal Prediction Filter Design (I)

Goal:

• **Find weight values to obtain good p filters (predictions) as a function of the video content→ low detail coefficients.**

Prediction model:



• Predict node k from spatial and temporal neighbors:

• Mean value of the spatial neighbors: $\bar{x}_s^{(k)} = \frac{1}{3}\left(x^{(2)} + x^{(3)} + x^{(5)}\right)$

• Mean value of the temporal neighbors: $\bar{x}_t^{(k)} = \frac{1}{2}\left(x^{(1)} + x^{(4)}\right)$

• Prediction of node k: $\hat{x}^{(k)} = w_s \bar{x}_s^{(k)} + w_t \bar{x}_t^{(k)}$

Optimal Prediction Filter Design (II)

Problem Formulation:

$$\min_{w_s, w_t} \sum_k \left(x^{(k)} - w_s \bar{x}_s^{(k)} - w_t \bar{x}_t^{(k)}\right)^2$$

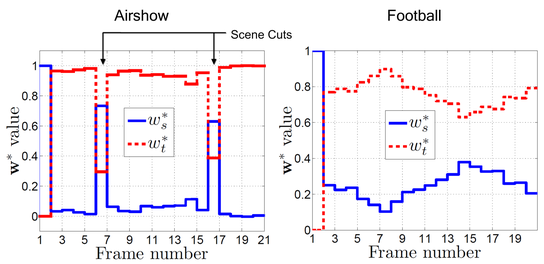
Problem Solution:

$$\mathbf{w}^* = (w_s^*, w_t^*) = \begin{bmatrix} \mathbf{x}\mathbf{A}_s\mathbf{A}_s^T\mathbf{x}^T & \mathbf{x}\mathbf{A}_s\mathbf{A}_t^T\mathbf{x}^T \\ \mathbf{x}\mathbf{A}_s\mathbf{A}_t^T\mathbf{x}^T & \mathbf{x}\mathbf{A}_t\mathbf{A}_t^T\mathbf{x}^T \end{bmatrix}^{-1} \cdot \begin{bmatrix} \mathbf{x}\mathbf{A}_s\mathbf{x}^T \\ \mathbf{x}\mathbf{A}_t\mathbf{x}^T \end{bmatrix}.$$

• A_s, A_t : spatial,temporal Adjacency matrices.
• **x:** data vector.

Results:

• Evolution of the optimal weights (calculated frame by frame).



• Mean energy per coefficient in the detail coefficients and j=1:

$$E_{d_{j=1}} = \frac{1}{|\mathcal{P}_{j=1}|} \sum_{m \in \mathcal{P}_{j=1}}^{|\mathcal{P}_{j=1}|} d_{m,j=1}^2$$

•Comparison against previous work [1], with fixed weights, w_t=10; w_s =2.

	$E_{d_{j=1}}$[1]	$E_{d_{j=1}}$ prop
Carphone	14	12
Mobile	44	37
Airshow (scene cut)	34	17
Football (fast motion)	408	240

Distributed U/P Assignment

Problem:

• U/P assignment complexity increases rapidly with the number of nodes *N*:

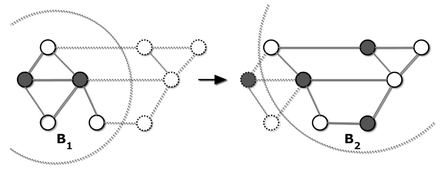
$$O\left(N^3 \cdot \log N\right)$$

Proposed solution:

• Work in a distributed manner.
• Complexity increases linearly with *N*:

$$O\left(\tfrac{N}{B} B^3 \cdot \log B\right)$$

▪B → block size used in the algorithm.
▪Complexity-precission trade-off in B.



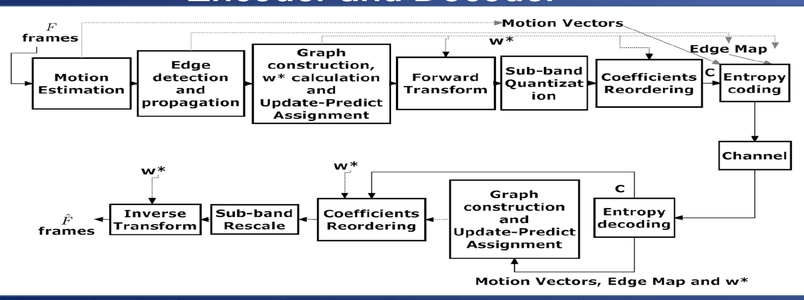
Results: $CR = \frac{time_{cent}}{time_{dist}}$

▪B=512

	Carphone	Mobile	Container
CR	228	197	203

▪Negligible loss in performance.

Encoder and Decoder



Experimental Results

Experimental Setup:

• **Quantization:** Subband dependent quantization.

• **Scanning:** Inter and Intra reordering [1].

• **Run length** encoding.

• **Arithmetic coding.**

• **5 levels** of the transform.

• Side info: Send the weight values every frame.

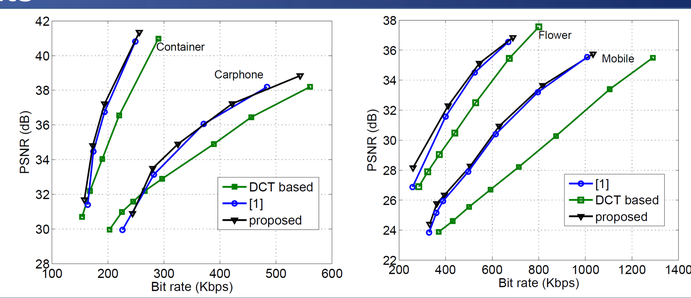
• **[1]→ wt=10; ws =2.**

• **B=512.**

Conclusions

• **Gain about 0.3-0.6 dB over [1], and 2-4 dB over DCT based encoder.**

• **Reduced complexity in comparison to [1].**



[1] E. Martínez-Enríquez, F. Díaz-de María, and A. Ortega, "Video encoder based on lifting transforms on graphs," in *Image Processing (ICIP), 2011 18th IEEE International Conference on, sept. 2011, pp. 3509 –3512.*